Pitch Shifting & Vocoding

musi4545
Fall 2021
Prof: Luke Dahl

Coming up…

Schedule for the rest of the semester: NOT ACCURATE…

- Remaining Topics: Pitch Shifting, Vocoders, Phasers, Reverb
- **Final Project:**
    - **Today:** Discuss ideas & questions in class
    - **Proposal** due Fri Nov 20, 11pm (on Collab).
    - **Progress Report:** Friday Nov 27 (send email)
- **Research & Writing** due Sat Nov 21, 11pm (on Collab)
- **Final Project Presentation** ~~Dec 7, 2-5pm (on Zoom)~~
    - Thursday December 3, 9:30am-noon (on Zoom)
- **Final Project Report** due Weds Dec 9, 11pm (on Collab)
- **Final Quiz** available Sun Nov 29, due Fri Dec 11, 11pm (on Collab)

Time-Domain Pitch Shifter

As we saw with the Chorus effect, slowly changing the read pointer of a delay line creates a pitch shift.

Continuously shortening the delay line raises the pitch. Continuously lengthening the delay lowers the pitch.

Think of the playing a tape faster or slower.

The problem is that if we continue to shift pitch in one direction we will go off one end of the delay line.

One solution is to cross-fade between two read pointers that are both moving the same direction and the same speed, but at different parts of the delay line.

Based on Zölzer DAFX 6.4.3 “Pitch Shifting by Delay-Line Modulation”

Time-Domain Pitch Shifter

We have seen that continually changing the read time from a delay line leads to a change in pitch. Moving the read time at a constant rate leads to a constant pitch shift. However we can't move the read time infinitely in either direction.

One solution/trick is to have multiple read pointers, spaced out along the delay line, all moving at the same rate. The output of each read is being faded in and out as it moves through the delay.

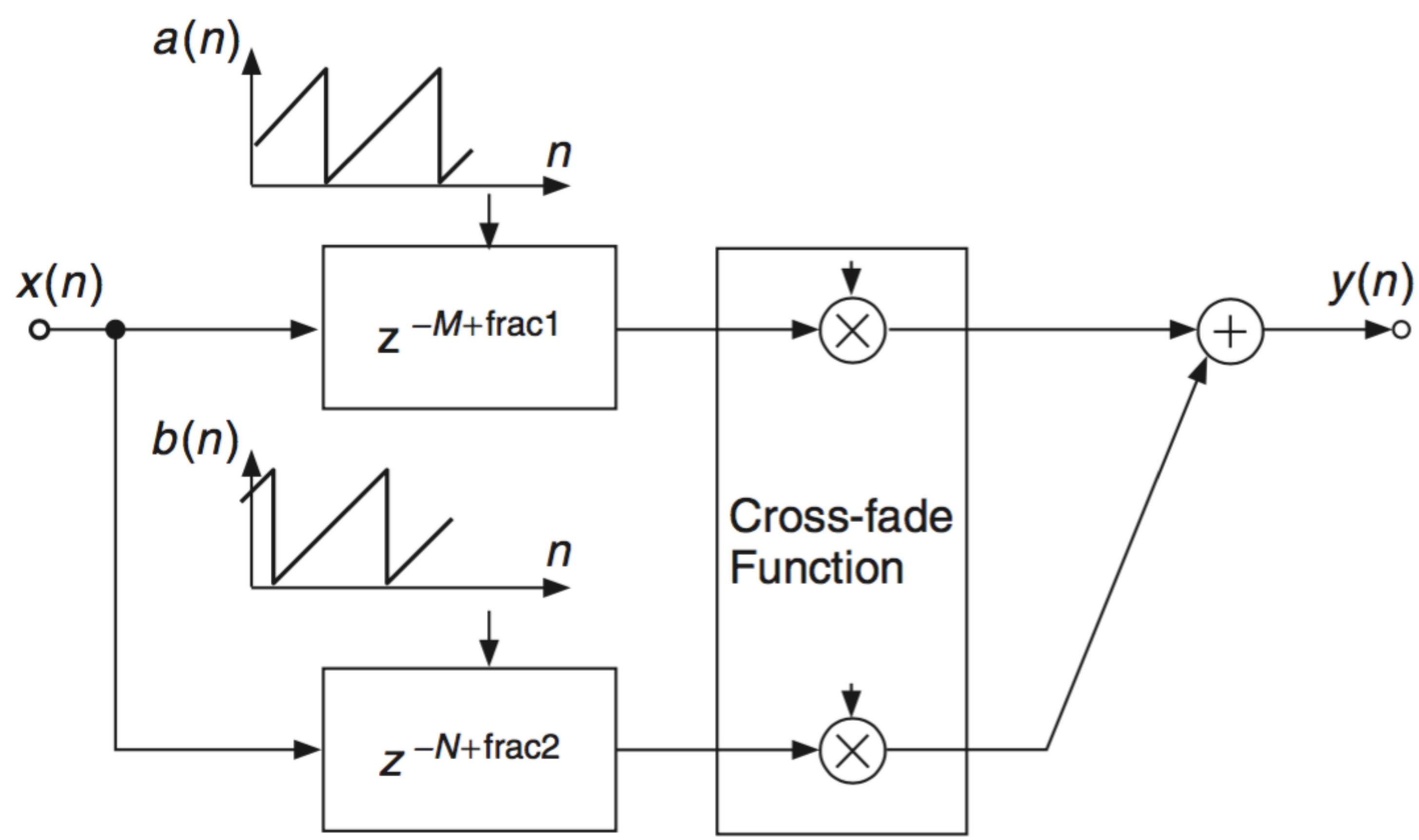


Figure 6.13 Pitch shifting.

A version with two delay lines (Zolzer)

Pitch Shifter (time-domain)

This is an instance of "overlap add" pitch-shifting. See Zolzer, *Digital Audio Effects*, 6.4.3 Pitch shifting by delay-line modulation.

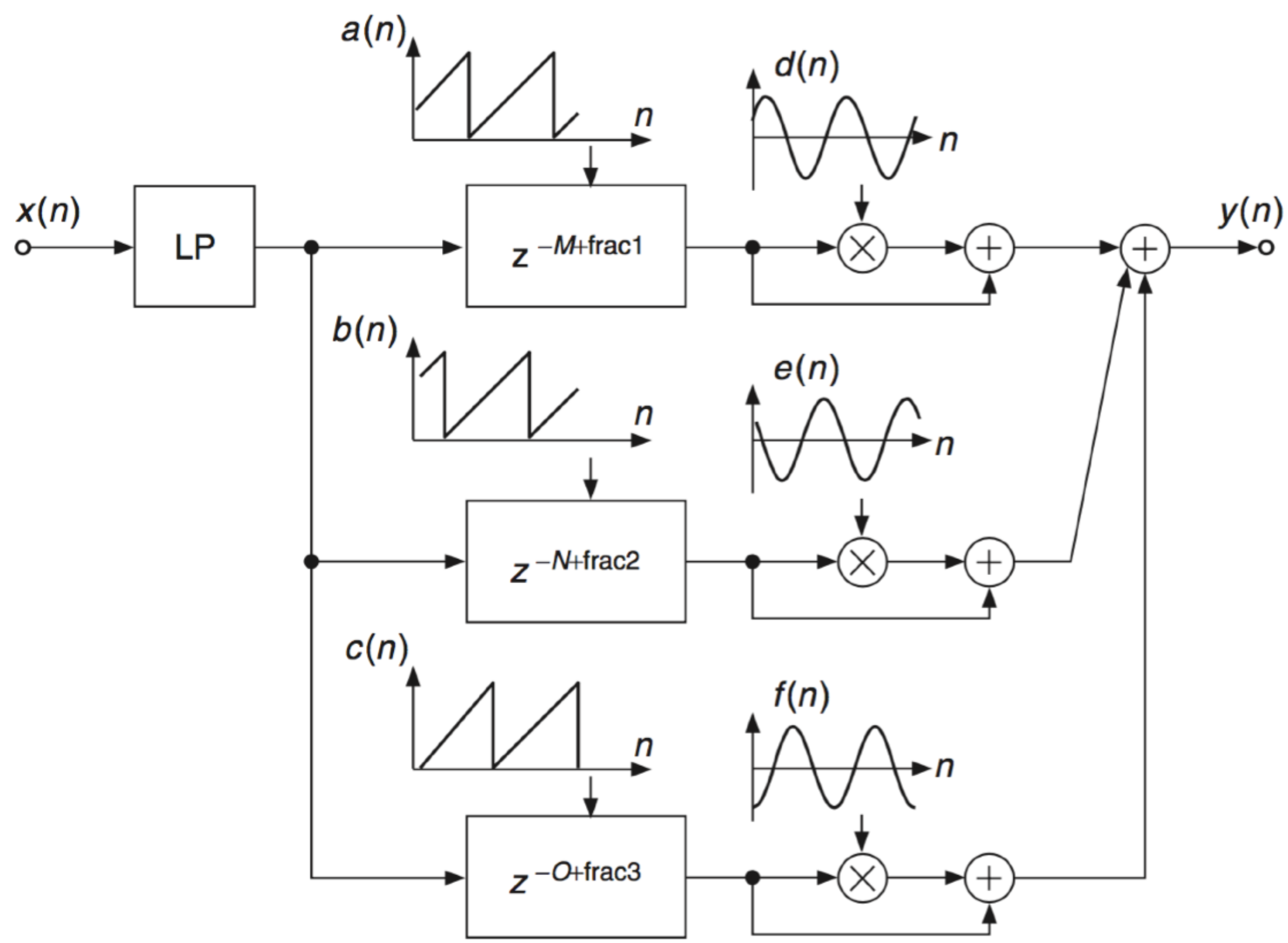


Figure 6.15 Enhanced pitch transposer: block diagram.

Zolzer'z time-domain pitch shifter that uses three delay lines

Pitch Shifter (time-domain)

We can use a raised-cosine window to control the amplitude of each read pointer.

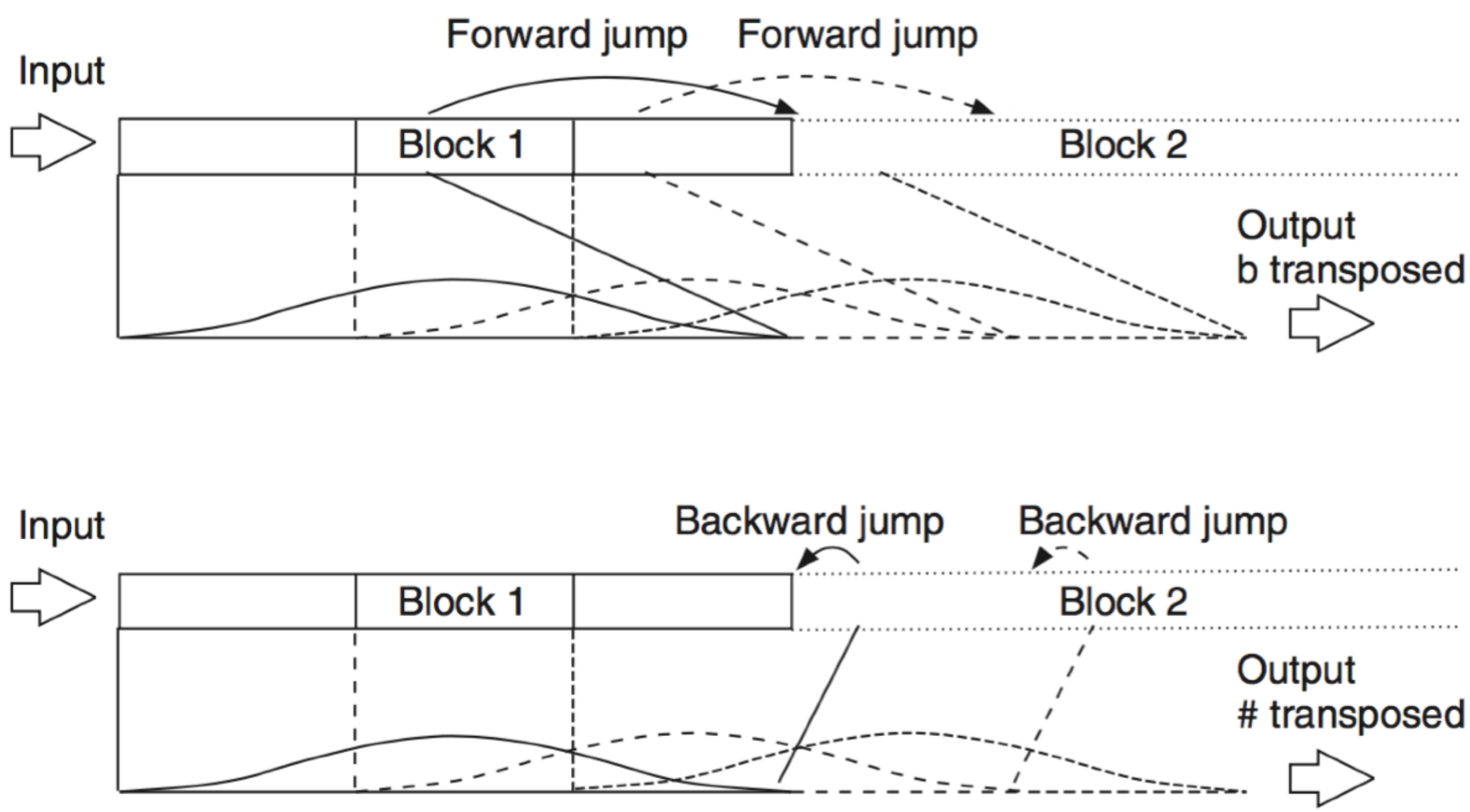
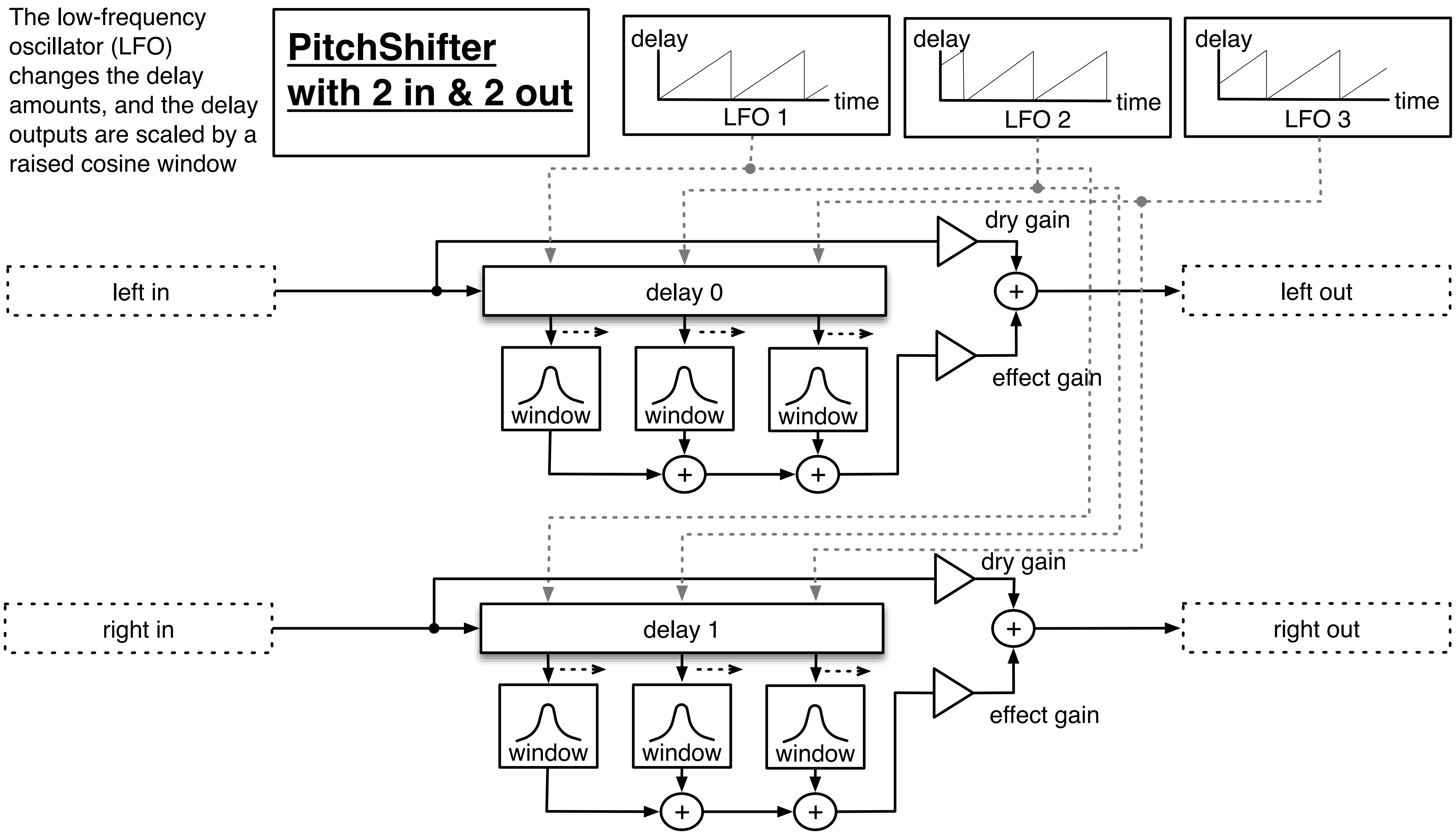


Figure 6.14 Enhanced pitch transposer: block processing, time shifting and overlap-add.

Zolzer's time-domain pitch shifter that uses three delay lines

Pitch Shifter (time-domain)

Here's my own diagram for the same algorithm:



Time-Domain Pitch Shifter

How do we change the delay time to create a desired pitch shift?

If we want to change the pitch by a number of semitones, the desired frequency ratio is:

$$Ratio = 2^{\left(\frac{semitones}{12}\right)}$$

The amount to increment the amount of delay (in samples) is:

```
delay_inc = 1 - Ratio;
```

So at each sample:

```
delay_in_samples += delay_inc;
```

Time-Domain Pitch Shifter

Pitch shifting and time compression/expansion are duals. It's difficult to change one without changing the other.

- Can raise the pitch by speeding up the sound
- Can speed up a sound but it raises the pitch.

Munchkinization: vowels don't sound right when pitch shifted.

Ways to do pitch shifting without changing time:

- Frequency domain techniques.
- Perform one of the following time compression/expansion techniques, then play back at different speed.

Ways to do time compression/expansion that break this constraint:

- Overlap Add (OLA)
- Synchronous Overlap Add (SOLA)
- Pitch Synchronous Overlap Add (PSOLA)

Analogue Pitch Shifter: The Phonogène

The same idea implemented on a tape machine by Pierre Schaeffer in ~1954

Analogue Pitch Shifter

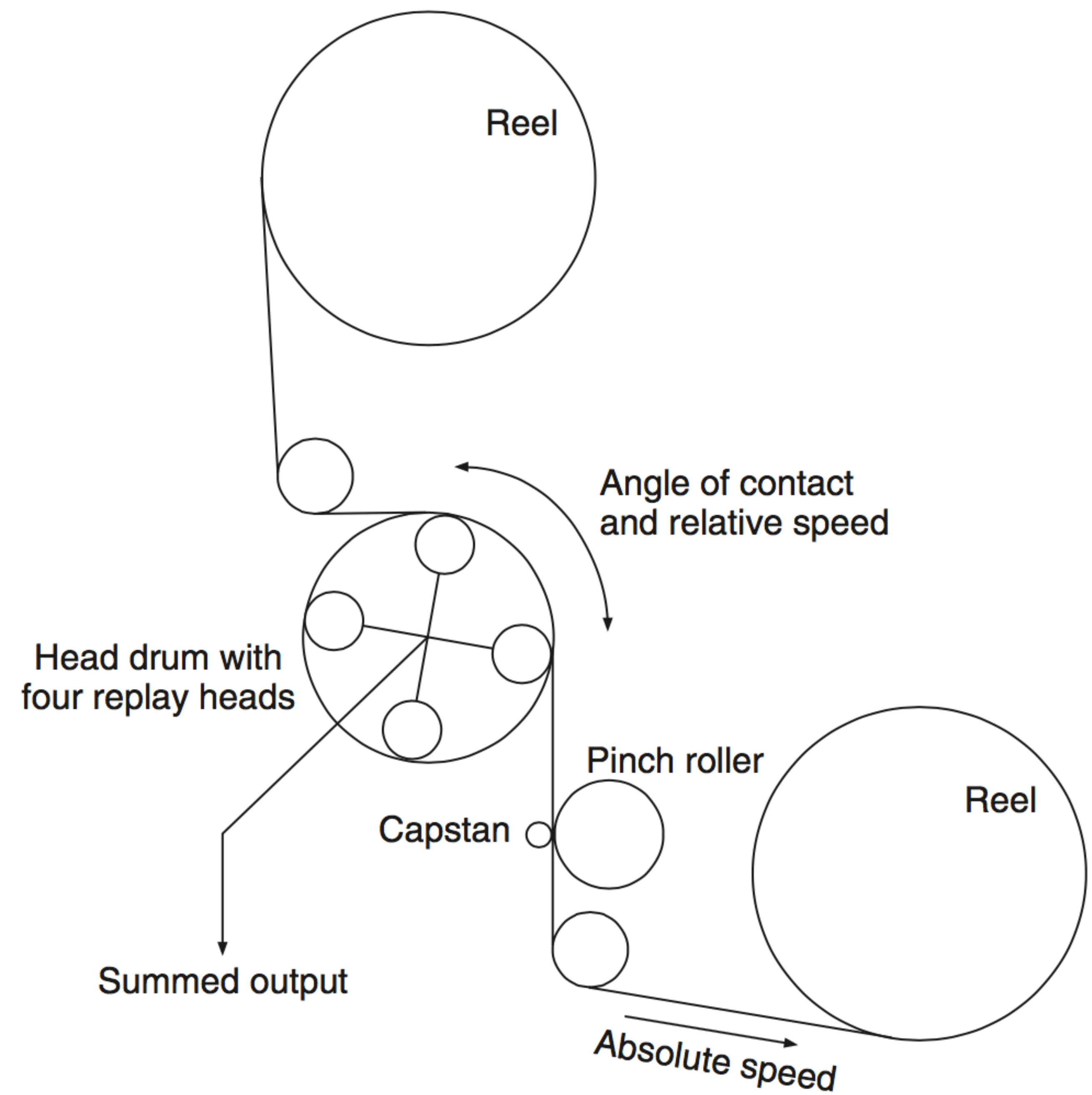


Figure 6.5 Tape-based time compression – expansion system, according to [Mol60].

Interlude: What makes different vowels sound different?

When you speak or sing a vowel, your vocal cords creates periodic short pulses at audio frequencies. This sound is “wideband” (i.e. is has energy across a wide range of frequencies.)

This sound resonates through the throat, mouth, nasal cavity, and lips. The shape of these acts as a filter, emphasizing some frequencies.

When you change the shape of your mouth, nasal cavity, and lips, the “filter” changes, emphasizing a different set of frequencies.

We can model this acoustic filter as a small number (2 to 5) of band-pass filters in parallel.

The pitch you sing determines what harmonics exist. The formants control how loud each harmonic is (they act as a **spectral envelope**)

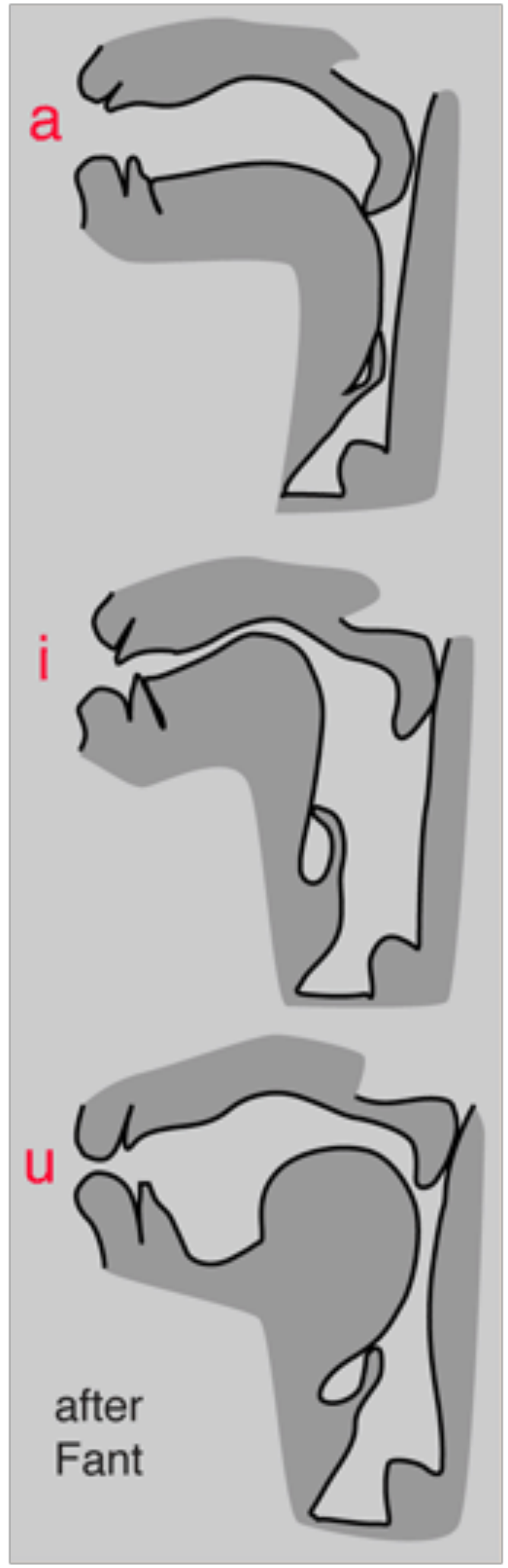
See: https://en.wikipedia.org/wiki/Formant

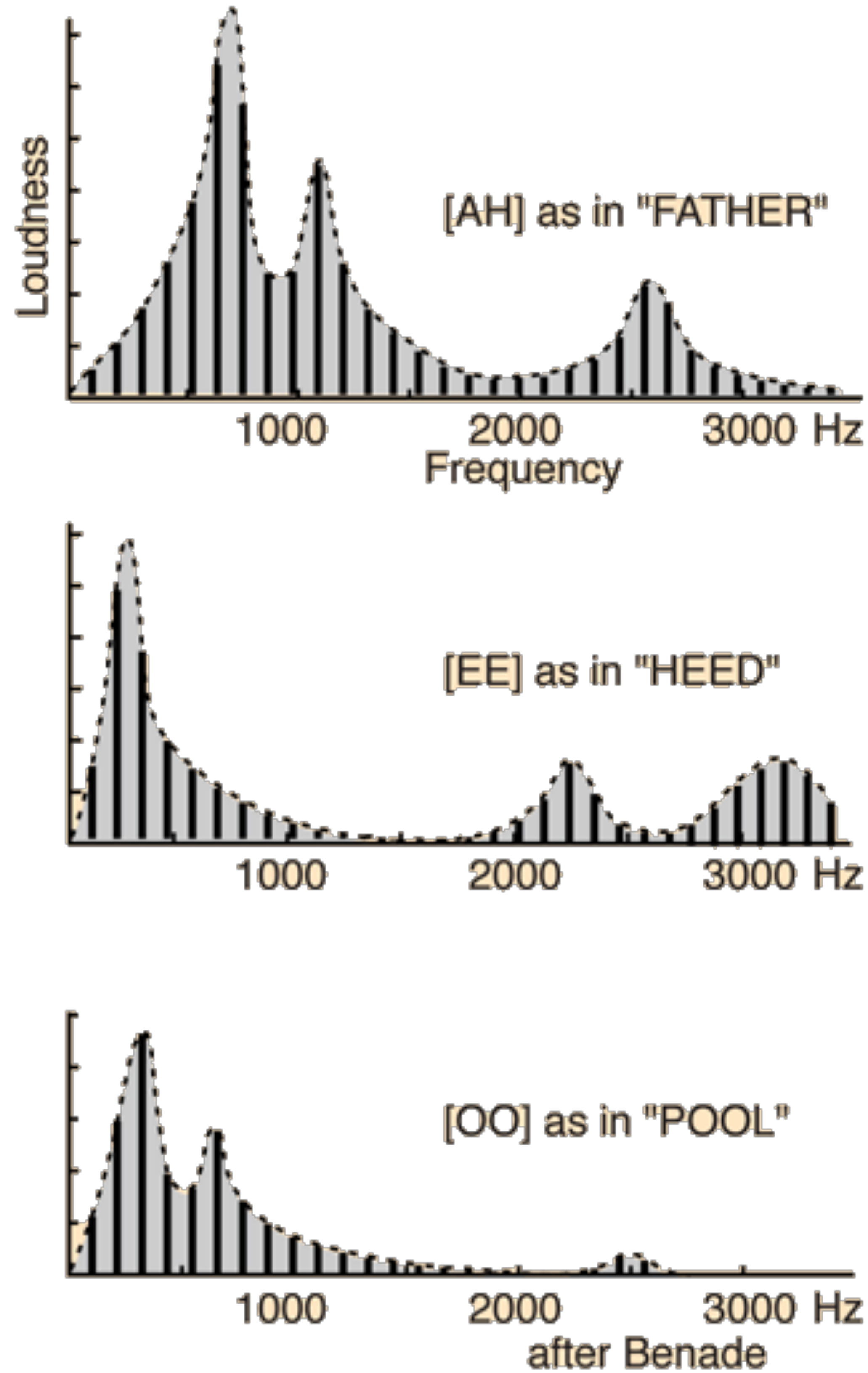
Why do vowels sound different?

The term “**formant**” can refer to the frequency of each resonance (or peak) in the frequency response of the filter created by the throat, mouth, nose, and lips.

In this diagram we see how the three peaks move to different frequencies for different vowels.

See: https://en.wikipedia.org/wiki/Formant





Source - Filter Model

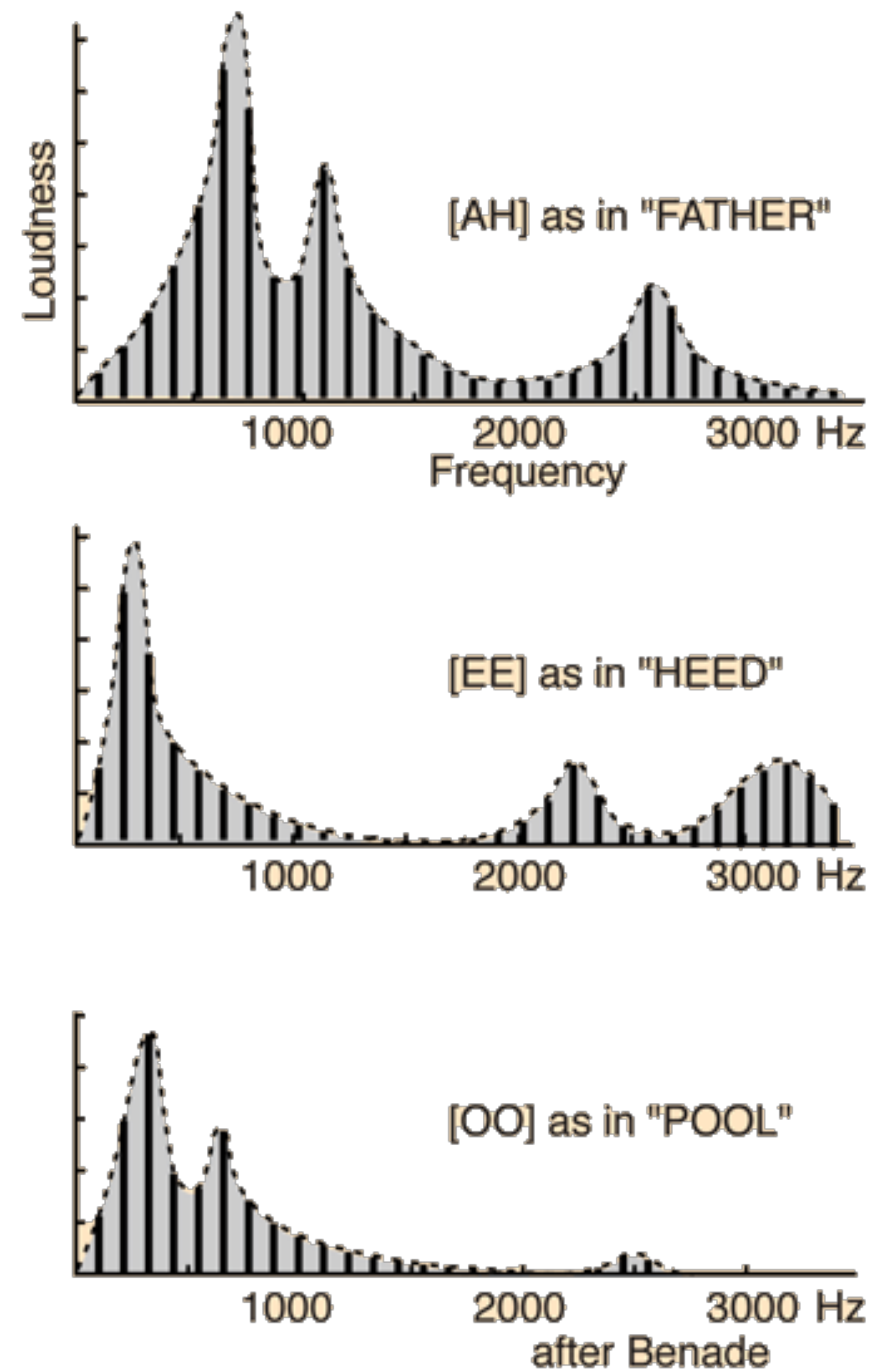
We can model the vocal production system as a source and a filter…

The **source** is the vocal chords, which emit short pulses at the rate of the pitch. (These pulses are *wideband*, i.e. having energy at a wide range of frequencies.)

The throat, mouth, lips, and nose function as a **filter**, to emphasize some frequencies and de-emphasize other…

…In other words it applies a **spectral envelope** to the source sound.

Other instruments can be modeled this way, e.g. string with resonant body.



Interlude: Talkbox

Talkbox is an “analog effect” where an instrument (typically an electric guitar) is made to talk.

Examples: http://ultimateclassicrock.com/talk-box-songs/

It works as follows:

- The guitar sound is played from a speaker which is attached to a flexible plastic tube.
- The guitarist places the tube inside their mouth, allowing the guitar sound to come out their mouth
- When the guitarist moves their mouth as if to speak, the mouth shapes the guitar sound making it sound like the vowels of speech.

Example: https://www.youtube.com/watch?v=ROtFteYkvoU (look for the clear tube running up the mic stand!)

Example: https://www.youtube.com/watch?v=L_CBZkd2tGE

Channel Vocoder

Vocal sounds are produced by the shape of the mouth, throat, and nasal cavity imposing a spectral envelope onto the pitched wide-band signal created by the vocal chords.

A vocoder is used to impress the spectral envelope of one sound (sound1) onto another sound (sound2).

For example if sound1 is someone singing, and sound2 is a wide-band pitched synthesizer sound, then the output will sound like the synthesizer is singing.

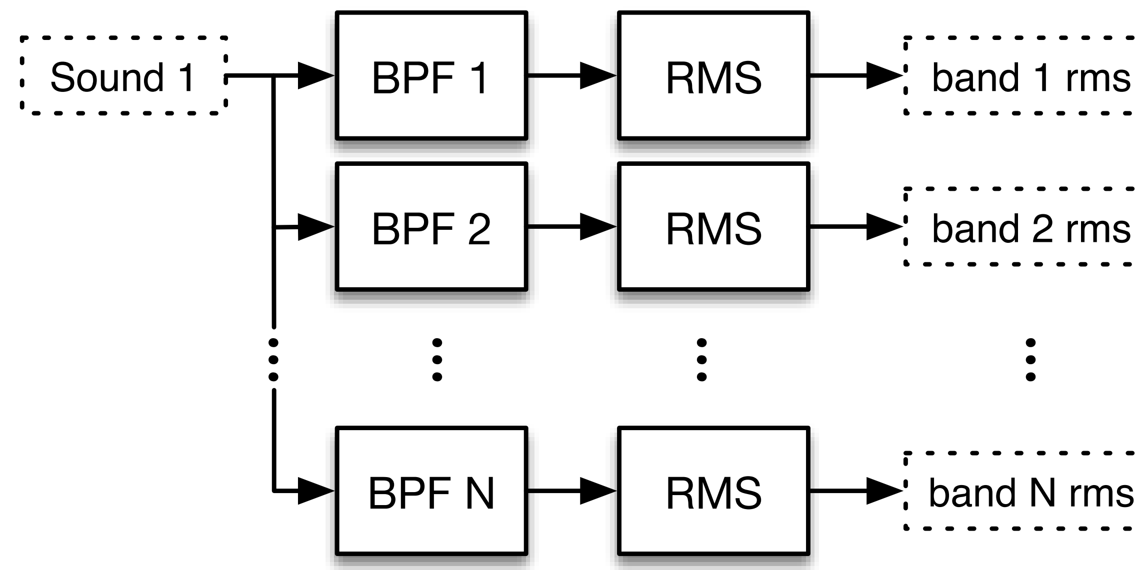
Channel Vocoder

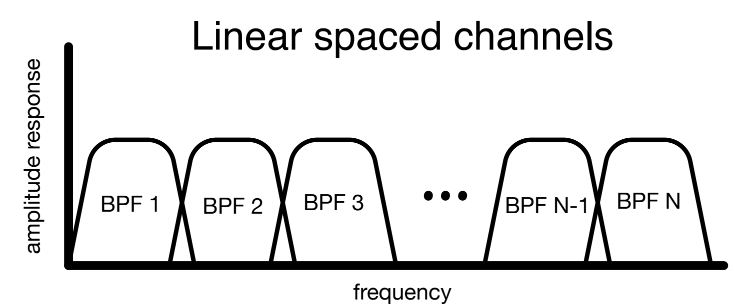
Sound1 is analyzed to extract its *spectral envelope.*

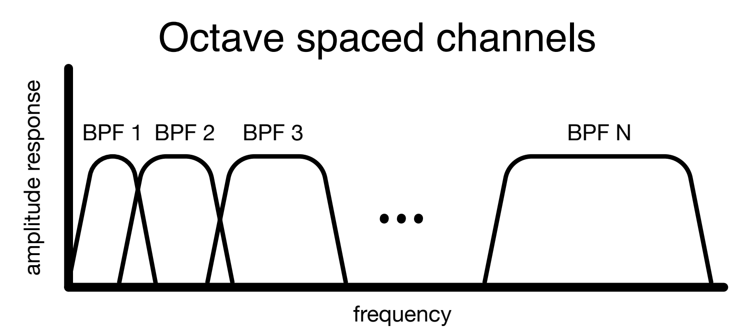
It is passed through a bank of N band-pass filters. These filters can be spaced linearly or logarithmically in frequency.

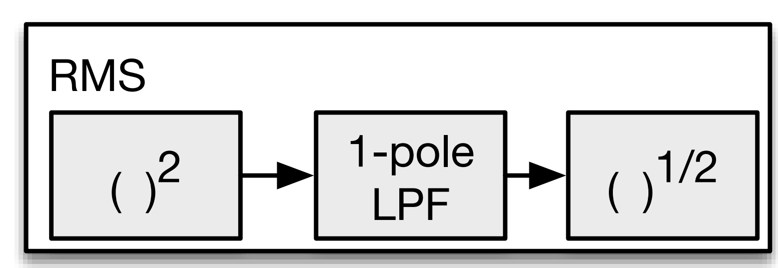
The RMS energy in each band is estimated.

The energy at each band approximates the spectral envelope.







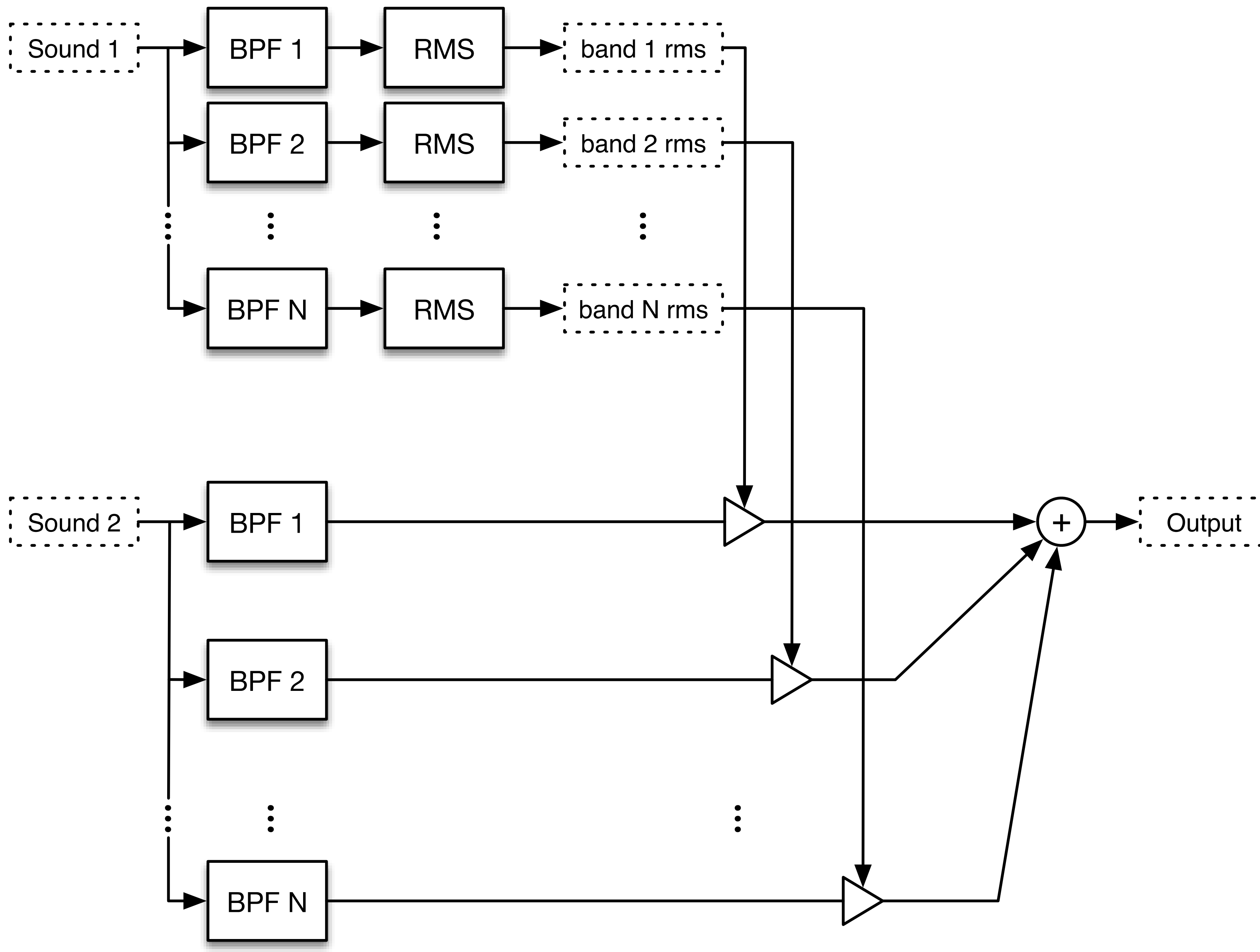


Channel Vocoder

Sound 2 is passed through an *identical* bank of band-pass filters.

The RMS energy level for each band in sound1 is used as a gain to the band-passed signals of sound2.

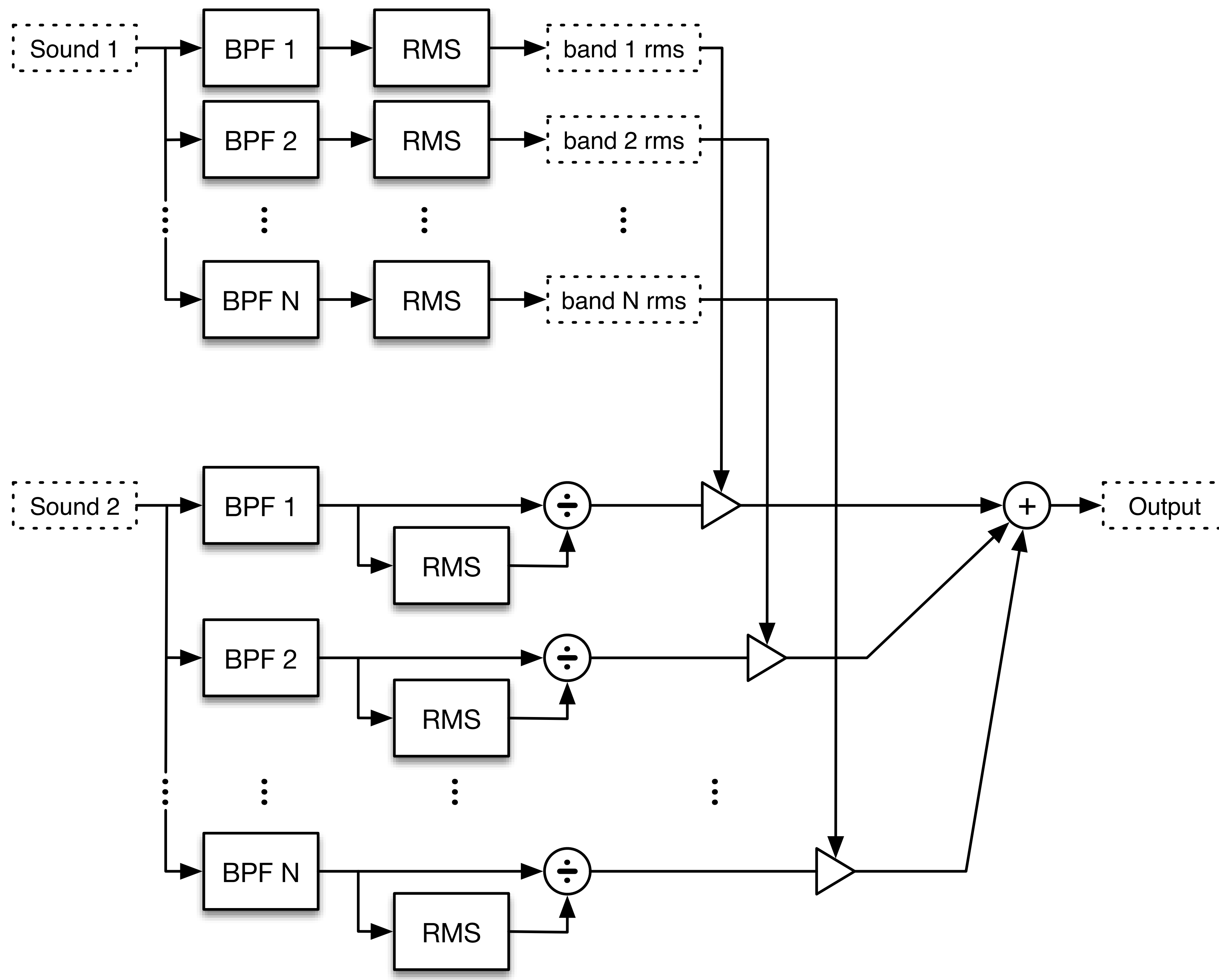
This imposes the spectral shape of sound1 onto sound2.



Channel Vocoder

This effect can be improved by **whitening** the spectrum of sound2 before applying the gains from sound1.

The RMS energy in each band of sound2 is calculated and then applied as an inverse gain (dividing) to the channel.



Pitch Shifter (time-domain)

Why does it sound like a munchkin or monster?

Formants…

Techniques for preserving formants…

- Pitch-synchronous Overlap Add (PSOLA)
- Frequency domain techniques that separate the spectral envelope from the spectrum.